Results on Parity-Check Matrices with Optimal Stopping and/or Dead-End Set Enumerators

Jos H. Weber, *Senior Member, IEEE,* and Khaled A.S. Abdel-Ghaffar, *Member, IEEE*

***Abstract*— The performance of iterative decoding techniques for linear block codes correcting erasures depends very much on the sizes of the stopping sets associated with the underlying Tanner graph, or, equivalently, the parity-check matrix representing the code. In this paper, we introduce the notion of dead-end sets to explicitly demonstrate this dependency. The choice of the parity-check matrix entails a trade-off between performance and complexity. We give bounds on the complexity of iterative decoders achieving optimal performance in terms of the sizes of the underlying parity-check matrices. Further, we fully characterize codes for which the optimal stopping set enumerator equals the weight enumerator.**

***Index Terms*— Dead-end set, iterative decoding, linear code, parity-check matrix, stopping set.**

I. Introduction

Iterative decoding techniques, especially when applied to low-density parity-check (LDPC) codes, have attracted a great attention recently. In these techniques, decoding is based on a Tanner graph determined by a parity-check matrix of the code, which does not necessarily, and typically does not, have full rank. It is well known that the performance of iterative decoding algorithms in case of binary erasure channels depends on the sizes of the stopping sets associated with the Tanner graph representing the code [3]. Several interesting results on stopping sets associated with Tanner graphs of given girths are given in [8], [11]. There are more specific results for classes of codes represented by particular Tanner graphs, see, e.g., [7], [1], [15], as well as more general results pertaining to ensembles of LDPC codes, see e.g., [2], [3], [6], [12].

In this paper, we define the notion of dead-end sets to explicitly show the dependency of the performance on the stopping sets. We then present several results that show how the choice of the parity-check matrix of the code, which determines decoding complexity, affects the stopping and the dead-end sets, which determine decoding performance. Our study differs from the aforementioned studies, but agrees with the studies by Schwartz and Vardy [13], Hollmann and Tolhuizen [5], Han and Siegel [4], and Weber and Abdel-Ghaffar [14], in its focus on the relationship between the stopping sets on one hand and the underlying code representation, rather than the code itself, on the other hand. Since linear algebra is used to study this relationship, for our purpose, parity-check matrices are more convenient than the equivalent Tanner graphs for code representation.

Let $\mathcal{C}$ be a binary linear $[n, k, d]$ block code, where n, k, and d denote the code's length, dimension, and Hamming distance, respectively. Such a code is a k-dimensional subspace of the space of binary vectors of length n, in which any two different elements differ in at least d positions. The set of codewords of $\mathcal{C}$ can be defined as the null space of the row space of an $r \times n$ binary parity-check matrix $\mathbf{H} = (h_{i,j})$ of rank $n-k$. Assuming all rows in $\mathbf{H}$ are different,

$$n - k \leq r \leq 2^{n-k}. \tag{1}$$

The row space of $\mathbf{H}$ is the $[n, n-k, d^{\perp}]$ dual code $\mathcal{C}^{\perp}$ of $\mathcal{C}$.

The support of a binary word $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is the set $\{j : x_j \neq 0\}$ and the weight of $\mathbf{x}$ is the size of its support. For the zero word $\mathbf{0} = (0, 0, \ldots, 0)$, the support is the empty set, $\emptyset$, and the weight is zero. Since a binary word $\mathbf{x}$ is a codeword of $\mathcal{C}$ if and only if $\mathbf{x}\mathbf{H}^{\mathrm{T}} = \mathbf{0}$, the parity-check matrix $\mathbf{H}$ gives rise to r parity-check equations, denoted by

$$\mathrm{PCE}_i(\mathbf{x}) : \textstyle\sum_{j=1}^{n} h_{i,j} x_j = 0 \text{ for } i = 1, 2, \ldots, r. \tag{2}$$

An equation $\mathrm{PCE}_i(\mathbf{x})$ is said to check $\mathbf{x}$ in position j if and only if $h_{i,j} = 1$.

On the binary erasure channel, each bit of the transmitted codeword is erased with probability ϵ, while it is received correctly with probability $1 - \epsilon$, where $0 < \epsilon < 1$. For a received word $\mathbf{r} = (r_1, r_2, \ldots, r_n)$, the erasure set is

$$\mathcal{E}_{\mathbf{r}} = \{j : r_j \neq 0, 1\}. \tag{3}$$

A received word can be decoded unambiguously if and only if it matches exactly one codeword of $\mathcal{C}$ on all its non-erased positions. Since $\mathcal{C}$ is a linear code, this is equivalent to the condition that the erasure set $\mathcal{E}_{\mathbf{r}}$ does not contain the support of a non-zero codeword. If $\mathcal{E}_{\mathbf{r}}$ does contain the support of a non-zero codeword, then it is said to be *incorrigible*. A decoder for $\mathcal{C}$ which achieves unambiguous decoding whenever the erased set is not incorrigible is said to be optimal for the binary erasure channel. An exhaustive decoder searching the complete set of codewords is optimal. However, such a decoder usually has a prohibitively high complexity.

Iterative decoding procedures may form a good alternative, achieving close to optimal performance at much lower complexity [9], in particular for LDPC codes. Here, we consider a well-known algorithm, often expressed in terms of a Tanner graph, which exploits the parity-check equations in order to determine the transmitted codeword. Initially, we set $\mathbf{c} = \mathbf{r}$ and

Manuscript submitted to IEEE Transactions on Information Theory, July 7, 2006. The authors were supported by the Dutch Ministry of Economic Affairs through the Airlink research project DTC.5961, and by the NSF through grants CCR-0117891 and ECS-0121469. This paper was presented in part at the International Symposium on Turbo Codes, Munich, Germany, April 3–7, 2006.

Jos H. Weber is with the Delft University of Technology, Faculty of EEMCS, Mekelweg 4, 2628 CD Delft, The Netherlands, j.h.weber@tudelft.nl. Khaled A.S. Abdel-Ghaffar is with the University of California, Department of ECE, Davis, CA 95616,USA, ghaffar@ece.ucdavis.edu.

$\mathcal{E}_{\mathbf{c}} = \mathcal{E}_{\mathbf{r}}$. If $\mathrm{PCE}_i(\mathbf{c})$ checks $\mathbf{c}$ in exactly one erased position j^*, then we use (2) to set

$$c_{j^*} = \sum_{j \notin \mathcal{E}_{\mathbf{c}}} h_{i,j} c_j \tag{4}$$

and we remove j^* from the erasure set $\mathcal{E}_{\mathbf{c}}$. Applying this procedure iteratively, the algorithm terminates if there is no parity-check equation left which checks exactly one erased symbol. Erasure sets for which this is the case have been named *stopping sets* [3]. In case the final erasure set $\mathcal{E}_{\mathbf{c}}$ is empty, the iterative algorithm retrieves all erased symbols, and thus the final word $\mathbf{c}$ is the transmitted codeword. In case the final erasure set $\mathcal{E}_{\mathbf{c}}$ is a non-empty stopping set, the iterative decoding process is unsuccessful. The final erasure set $\mathcal{E}_{\mathbf{c}}$ is the union of the stopping sets contained in $\mathcal{E}_{\mathbf{r}}$, and thus $\mathcal{E}_{\mathbf{c}}$ is empty if and only if $\mathcal{E}_{\mathbf{r}}$ contains no non-empty stopping set. Therefore, we introduce the notion of a *dead-end set* for an erasure set which contains at least one non-empty stopping set. In summary, on the binary erasure channel, an optimal decoder is unsuccessful if and only if $\mathcal{E}_{\mathbf{r}}$ is an incorrigible set, and an iterative decoder is unsuccessful if and only if $\mathcal{E}_{\mathbf{r}}$ is a dead-end set.

This paper is organized as follows. In Section II we characterize codeword supports, incorrigible sets, stopping sets, and dead-end sets in terms of a parity-check matrix and derive basic results from this characterization. We also review results from [13] and [5] which are most relevant to this work. Dead-end sets and stopping sets are studied in Sections III and IV, respectively. Conclusions are presented in Section V.

II. Definitions and Preliminaries

Again, let $\mathcal{C}$ be a linear binary $[n, k, d]$ block code with an $r \times n$ binary parity-check matrix $\mathbf{H} = (h_{i,j})$ of rank $n-k$. Let $\mathcal{S}$ be a subset of $\{1, 2, \ldots, n\}$. For any $\mathbf{H}$ and $\mathcal{S}$, let $\mathbf{H}_{\mathcal{S}}$ denote the $r \times |\mathcal{S}|$ submatrix of $\mathbf{H}$ consisting of the columns indexed by $\mathcal{S}$. A set $\mathcal{S}$ is the support of a codeword if and only if all rows in $\mathbf{H}_{\mathcal{S}}$ have even weight, i.e., if and only if

$$|\{j \in \mathcal{S} : h_{i,j} = 1\}| \equiv 0(2) \quad \forall i = 1, 2, \ldots, r. \tag{5}$$

A set $\mathcal{S}$ is an incorrigible set if and only if it contains the support of a non-zero codeword. A set $\mathcal{S}$ is a stopping set for the parity-check matrix $\mathbf{H}$ if and only if $\mathbf{H}_{\mathcal{S}}$ does not contain a row of weight one, i.e., if and only if

$$|\{j \in \mathcal{S} : h_{i,j} = 1\}| \neq 1 \quad \forall i = 1, 2, \ldots, r. \tag{6}$$

Hence, the support of any codeword is a stopping set. A set $\mathcal{S}$ is a dead-end set for the parity-check matrix $\mathbf{H}$ if and only if it contains a non-empty stopping set.

The polynomial $A(x) = \sum_{i=0}^{n} A_i x^i$, where A_i is the number of codewords of weight i, is called the *weight enumerator* of code $\mathcal{C}$. Similarly, $I(x) = \sum_{i=0}^{n} I_i x^i$, where I_i is the number of incorrigible sets of size i, is called the *incorrigible set enumerator* of $\mathcal{C}$. Clearly,

$$d = \min\{i \geq 1 : A_i > 0\} = \min\{i \geq 0 : I_i > 0\} \tag{7}$$

and

$$A_i = \begin{cases} 1 & \text{if } i = 0, \\ 0 & \text{if } 1 \leq i \leq d-1. \end{cases} \tag{8}$$

The incorrigible set enumerator satisfies

$$I_i = \begin{cases} 0 & \text{if } 0 \leq i \leq d-1, \\ A_i & \text{if } i = d, \\ \binom{n}{i} & \text{if } n-k+1 \leq i \leq n, \end{cases} \tag{9}$$

where the last property follows from the observation that any set $\mathcal{S}$ of size $|\mathcal{S}| > n-k$ contains the support of a non-zero codeword as the rank of $\mathbf{H}_{\mathcal{S}}$ is at most $n-k$.

The polynomials $S(x) = \sum_{i=0}^{n} S_i x^i$, where S_i is the number of stopping sets of size i, and $D(x) = \sum_{i=0}^{n} D_i x^i$, where D_i is the number of dead-end sets of size i, are called the *stopping set enumerator* and the *dead-end set enumerator*, respectively, of parity-check matrix $\mathbf{H}$. From the observation that (5) and (6) are equivalent for sets $\mathcal{S}$ with $|\mathcal{S}| \leq 2$, it follows that

$$S_i = A_i \text{ and } D_i = I_i \text{ if } 0 \leq i \leq 2. \tag{10}$$

In particular, $S_0 = 1$, $S_1 = S_2 = 0$, and $D_0 = D_1 = D_2 = 0$ for any parity-check matrix of a code of minimum distance $d \geq 3$.

Let s denote the smallest size of a non-empty stopping set (and thus the smallest size of a dead-end set), i.e.,

$$s = \min\{i \geq 1 : S_i > 0\} = \min\{i \geq 0 : D_i > 0\}. \tag{11}$$

The number s is called the stopping distance for the parity-check matrix $\mathbf{H}$ in [13]. For any parity-check matrix $\mathbf{H}$ of a binary linear $[n, k, d]$ block code $\mathcal{C}$, it holds that the stopping set enumerator satisfies

$$S_i = \begin{cases} 1 & \text{if } i = 0, \\ 0 & \text{if } 1 \leq i \leq s-1, \\ \binom{n}{i} & \text{if } n - d^{\perp} + 2 \leq i \leq n. \end{cases} \tag{12}$$

where the first property follows from (10) and (8), the second property follows from the definition of s, and the third property follows from the fact that the weight of any row in $\mathbf{H}_{\mathcal{S}}$ is either 0 or at least equal to $d^{\perp} - (d^{\perp} - 2) = 2$ for any $\mathcal{S}$ with $|\mathcal{S}| \geq n - d^{\perp} + 2$. Further, again for any parity-check matrix $\mathbf{H}$, it follows from the definitions of the various enumerators, (9), and (12), that the dead-end set enumerator satisfies

$$D_i = \begin{cases} 0 & \text{if } 0 \leq i \leq s-1, \\ S_i & \text{if } i = s, \\ \binom{n}{i} & \text{if } n-k+1 \leq i \leq n. \end{cases} \tag{13}$$

For code $\mathcal{C}$ on the binary erasure channel, the probability of unsuccessful decoding (UD) for an optimal (OPT) decoder is

$$P_{\mathrm{UD}}^{\mathrm{OPT}}(\mathcal{C}) = \sum_{i=d}^{n} I_i \epsilon^i (1-\epsilon)^{n-i} \sim I_d \epsilon^d = A_d \epsilon^d. \tag{14}$$

Similarly, the probability of unsuccessful decoding for an iterative (IT) decoder based on parity-check matrix $\mathbf{H}$ is

$$P_{\mathrm{UD}}^{\mathrm{IT}}(\mathbf{H}) = \sum_{i=s}^{n} D_i \epsilon^i (1-\epsilon)^{n-i} \sim D_s \epsilon^s = S_s \epsilon^s. \tag{15}$$

Hence, these two probabilities are completely determined by the incorrigible and dead-end set enumerators. Notice from (14) and (15) that iterative decoding is optimal if and only if $D(x) = I(x)$. At small erasure probabilities, $P_{\mathrm{UD}}^{\mathrm{OPT}}(\mathcal{C})$

and $P_{\mathrm{UD}}^{\mathrm{IT}}(\mathbf{H})$ are dominated by the terms $A_d\epsilon^d$ and $S_s\epsilon^s$, respectively. Actually, for sufficiently small values of ϵ, the parameters d and s are the most important parameters characterizing the performance of optimal decoding and iterative decoding, respectively. In (10) it is stated that if $i \leq 2$, then $S_i = A_i$. Therefore, $s = d$ for any parity-check matrix $\mathbf{H}$ of a code with $d \leq 3$, which is derived as Theorem 3 in [13]. Here, we show that this cannot be extended further.

Theorem 1: For any code $\mathcal{C}$ with Hamming distance $d \geq 4$, there exists a parity-check matrix $\mathbf{H}$ for which $s = 3$.

Proof: We may order the positions so that $\mathcal{C}$ has a codeword composed of d ones followed by $n-d$ zeros. In particular, the first d columns in any given parity-check matrix of $\mathcal{C}$ are linearly dependent, but no $d-1$ columns are such. The row space of the submatrix composed of these first d columns has dimension $d-1$ and a sequence of length d belongs to this row space if and only if its weight is even. By elementary row operations, we can obtain a parity-check matrix of the form

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}'_d & \mathbf{H}' \\ \mathbf{0} & \mathbf{H}'' \end{pmatrix}, \tag{16}$$

for some matrices $\mathbf{H}'$ and $\mathbf{H}''$ of appropriate sizes, where $\mathbf{H}'_d$ is the $(d-1) \times d$ matrix given by

$$\mathbf{H}'_d = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 \end{pmatrix}. \tag{17}$$

Clearly, $\mathcal{S} = \{1, 2, 3\}$ is a stopping set for $\mathbf{H}$ as no row of $\mathbf{H}_{\mathcal{S}}$ has weight one. ■

Contrary to the weight enumerator and the incorrigible set enumerator, which are fixed for a code $\mathcal{C}$, the stopping and dead-end set enumerators depend on the choice of the parity-check matrix $\mathbf{H}$. Theorem 1 shows that no matter how large the Hamming distance of the code is, a bad choice of the parity-check matrix may lead to very poor performance. Therefore, it is important to properly select the parity-check matrix of a code when applying iterative decoding.

Clearly, adding rows to a parity-check matrix does not increase any coefficient of the stopping set enumerator or the dead-end set enumerator. On the contrary, these coefficients may actually decrease at the expense of higher decoding complexity. The rows to be added should be in the dual code $\mathcal{C}^{\perp}$ of $\mathcal{C}$. By having all 2^{n-k} codewords in $\mathcal{C}^{\perp}$ as rows, we obtain a parity-check matrix that gives the best possible performance, but also the highest complexity, when applying iterative decoding. Since the order of the rows does not affect the decoding result, we refer to such matrix, with some ordering imposed on its rows which is irrelevant to our work, as the complete parity-check matrix of the code $\mathcal{C}$, and denote it by $\mathbf{H}^{\star}$. Its stopping set enumerator is denoted by $S^{\star}(x) = \sum_{i=0}^{n} S_i^{\star} x^i$, its dead-end set enumerator by $D^{\star}(x) = \sum_{i=0}^{n} D_i^{\star} x^i$, and its stopping distance by $s^{\star}$. Since the support of any codeword is a stopping set for any parity-check matrix, we have

$$S_i \geq S_i^{\star} \geq A_i \text{ and } D_i \geq D_i^{\star} \geq I_i \;\forall i = 0, 1, \ldots, n. \tag{18}$$

Consequently, $s \leq s^{\star} \leq d$, and $S^{\star}(x)$ and $D^{\star}(x)$ are called the code's optimal stopping set enumerator and optimal dead-end set enumerator, respectively. Schwartz and Vardy [13] have shown that

$$s^{\star} = d \tag{19}$$

and the results derived recently by Hollmann and Tolhuizen [5] imply, in addition, that

$$S_d^{\star} = A_d \tag{20}$$

and

$$D^{\star}(x) = I(x). \tag{21}$$

Actually, Schwartz and Vardy [13] have shown that, for $d \geq 3$, it is possible to construct a parity-check matrix with at most $\sum_{i=1}^{d-2} \binom{n-k}{i}$ rows for which $s = d$. They also obtain interesting results on the minimum number of rows in a parity-check matrix for which $s = d$. They obtain general bounds on this minimum number, which they call the stopping redundancy, as well as bounds for specific codes such as the Golay code and Reed-Muller codes. Han and Siegel [4] derived another general upper bound on the stopping redundancy for $d \geq 2$ given by $\sum_{i=1}^{\lceil (d-1)/2 \rceil} \binom{n-k}{2i-1}$.

Hollmann and Tolhuizen [5] specified rows that can be formed from any $(n-k) \times n$ parity-check matrix of rank $n-k$ to yield a parity-check matrix for which $D_i = I_i$ for $0 \leq i \leq m$, where m is any given integer such that $2 \leq m \leq n-k$. They have shown that the number of rows in the smallest parity-check matrix achieving this is at most $\sum_{i=0}^{m-1} \binom{n-k-1}{i}$.

Example 1: Let $\mathcal{C}$ be the $[8, 4, 4]$ Reed-Muller code. One of the parity-check matrices of $\mathcal{C}$ is

$$\mathbf{H}_8 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}. \tag{22}$$

For $i = 4, 5, 6, 7$, deleting the last $8-i$ rows in $\mathbf{H}_8$ still gives a parity-check matrix $\mathbf{H}_i$ for the code $\mathcal{C}$. Table I gives the stopping set enumerator $S(x)$ and the dead-end set enumerator $D(x)$ for the parity-check matrix $\mathbf{H}_i$ for $i = 4$, 5, and 8. The table also gives $S^{\star}(x)$ and $D^{\star}(x)$ corresponding to the complete parity-check matrix $\mathbf{H}^{\star}$, and the weight enumerator $A(x)$ and the incorrigible set enumerator $I(x)$. We point out that the matrix $\mathbf{H}_4$ is a frequently used full rank matrix for this code. For this matrix $s = 3$. The matrix $\mathbf{H}_5$ is the matrix proposed by Schwartz and Vardy [13] to achieve $s = 4$. For this matrix $S_4 > A_4$. The matrix $\mathbf{H}_8$ is constructed based on the techniques proposed by Hollmann and Tolhuizen to achieve $S_4 = A_4$. For later purposes, we also define the matrix $\mathbf{H}_{14}$ whose rows are the fourteen non-zero codewords in $\mathcal{C}^{\perp} =$

$\mathcal{C}$ of weight four. The stopping set enumerator and the dead-end set enumerator for this parity-check matrix are also listed in the table.

III. Dead-End Set Results

In this section, we investigate parity-check matrices for which the iterative decoding procedure achieves optimal performance, i.e., for which

$$P_{\mathrm{UD}}^{\mathrm{IT}}(\mathbf{H}) = P_{\mathrm{UD}}^{\mathrm{OPT}}(\mathcal{C}). \tag{23}$$

In order to satisfy (23), it is necessary and sufficient that the dead-end set enumerator equals the incorrigible set enumerator, i.e., $D(x) = I(x)$. From (21), we know that this is the case for the complete parity-check matrix, which contains 2^{n-k} rows. However, from a decoding complexity point of view, it may be desirable or required to reduce the number of rows in the parity-check matrix. Hence, an interesting research challenge is to find a parity check matrix $\mathbf{H}$ for code $\mathcal{C}$, with a minimum number of rows, but still having $D(x) = I(x)$.

As stated before, it is shown in [5] that there exists a parity-check matrix $\mathbf{H}$ with at most $\sum_{i=0}^{m-1} \binom{n-k-1}{i}$ rows for which $D_i = I_i$, $0 \le i \le m$, for any $1 \le m \le n-k$. By taking $m = n-k$ and noticing that $D_i = I_i$ for $n-k+1 \le i \le n$ from (9) and (13), we deduce the following result.

Theorem 2 (Hollmann and Tolhuizen): Let $\mathcal{C}$ be an $[n,k,d]$ binary linear code with $k < n$. Then, there exists a parity-check matrix with at most 2^{n-k-1} rows for which $D(x) = I(x)$.

Hollmann and Tolhuizen also show that for some codes, and in particular for Hamming codes, $D(x) \neq I(x)$ for any parity-check matrix with less than 2^{n-k-1} rows. However, depending on the code, it may be possible to reduce the number of rows in a parity-check matrix for which $D(x) = I(x)$ below 2^{n-k-1} as we show next.

Theorem 3: Let $\mathbf{H}$ be the matrix whose rows are the non-zero codewords in $\mathcal{C}^\perp$ of weight at most $k+1$. Then, $\mathbf{H}$ is a parity-check matrix for $\mathcal{C}$ and for this matrix $D(x) = I(x)$.

Proof: Let $\mathbf{H}'$ be an $(n-k) \times n$ parity-check matrix for the code $\mathcal{C}$. Then, there is a subset $\mathcal{S}$ of $\{1,2,\ldots,n\}$ of size $n-k$ such that $\mathbf{H}'_{\mathcal{S}}$ is an $(n-k) \times (n-k)$ matrix of rank $n-k$. The row space of this matrix contains every unit weight vector of length $n-k$. Therefore, the row space of $\mathbf{H}'$ contains $n-k$ vectors such that each vector has exactly a single one in a unique position indexed by an element in $\mathcal{S}$. Since these vectors have weight at most $k+1$ and are linearly independent, it follows that $\mathbf{H}$, which contains all of them as rows, has rank $n-k$ and is indeed a parity-check matrix for $\mathcal{C}$.

Next, we prove that for this matrix $D(x) = I(x)$, i.e., $D_i = I_i$ for $i = 0,1,\ldots,n$. From (10), (9), and (13), it suffices to show that $D_i = I_i$ for $3 \le i \le n-k$. For such an i, assume that $\mathcal{S}'$ is a subset of $\{1,2,\ldots,n\}$ of size i which does not contain the support of a non-zero codeword. Then, the columns of the $(n-k) \times n$ parity-check matrix $\mathbf{H}'$ indexed by the elements in $\mathcal{S}'$ are linearly independent. As $\mathbf{H}'$ has rank $n-k$, there is a set $\mathcal{S}''$ such that $\mathcal{S}' \subseteq \mathcal{S}'' \subseteq \{1,2,\ldots,n\}$ and $\mathbf{H}'_{\mathcal{S}''}$ is an $(n-k) \times (n-k)$ matrix of rank $n-k$. From the argument given in the first part of this proof, $\mathbf{H}$ contains $n-k$ vectors such that each vector has exactly a single one in a unique position indexed by an element in $\mathcal{S}''$, and in particular each vector has weight at most $k+1$. The existence of any one of the i vectors with a single one in a position indexed by an element in $\mathcal{S}'$ proves that $\mathcal{S}'$ is not a stopping set for $\mathbf{H}$. We conclude that every stopping set of size i for $\mathbf{H}$ contains the support of a non-zero codeword. Hence, $D_i = I_i$ for all i. ■

Let $H(x)$ denote the well-known binary entropy function $-x\log_2 x - (1-x)\log_2(1-x)$ for $0 < x < 1$.

Theorem 4: Let $\mathcal{C}$ be an $[n,k,d]$ binary linear code with $k \le n/2 - 1$. Then, there exists a parity-check matrix with at most $2^{nH((k+1)/n)}$ rows for which $D(x) = I(x)$.

Proof: From the bounds on the sum of binomial coefficients as presented on page 310 of [10], it follows that the number of codewords in the dual code of weight less than or equal to $k+1$ is at most equal to $2^{nH((k+1)/n)}$. Hence, the result follows from Theorem 3. ■

Note that the bound from Theorem 4 improves upon the bound from Theorem 2 for low-rate codes.

IV. Stopping Set Results

As stated earlier, iterative decoding based on a parity-check matrix is optimal, in the sense of having the smallest possible unsuccessful decoding probability on the binary erasure channel, if and only if $D(x)$ for this matrix is identical to $I(x)$ for code $\mathcal{C}$. This holds for the complete parity-check matrix as well as other matrices, whose sizes are bounded in Section III. For $D(x)$ to be identical to $I(x)$, we should have $s = d$ and $S_d = A_d$. Table I shows that it is possible to achieve optimal decoding using parity-check matrices, such as $\mathbf{H}_8$, with much smaller number of rows than in the complete parity-check matrix $\mathbf{H}^\star$. This is true in spite of the fact that these smaller matrices have stopping set enumerators that are different from $S^\star(x)$. We may wonder then what is the effect, if any, of the stopping set coefficients S_i for $i > d$ on performance. Notice that in this paper we defined the probability of unsuccessful decoding as the probability that the decoder fails to retrieve the transmitted codeword. Although an iterative decoder is unsuccessful in case the erasure set is a dead-end set, it still succeeds in retrieving those erased bits whose indices do not belong to any of the stopping sets contained in the erasure set. Therefore, it may be desirable to choose parity-check matrices for which $S_i = A_i$ not only for $i = d$ but also for $i > d$. Since $S^\star(x) \neq A(x)$ in Example 1, it follows that this is not possible in general. In fact, Theorem 6 will show that $S^\star(x) = A(x)$ only for a rather degenerate class of codes. Hence, the best that we may hope for is to have parity-check matrices, smaller than the complete parity-check matrix to reduce complexity, for which $S(x) = S^\star(x)$. The matrix $\mathbf{H}_{14}$, specified in Theorem 3, is one such matrix for the $[8,4,4]$ Reed-Muller code. Actually, it can be checked that this is the smallest parity-check matrix for this code satisfying $S(x) = S^\star(x)$.

Example 2: Let $\mathcal{C}$ be the $[8,4,4]$ Reed-Muller code considered in Example 1. From Table I, we notice that the iterative decoders based on $\mathbf{H}^\star$, $\mathbf{H}_{14}$, and $\mathbf{H}_8$ achieve the smallest possible probability of unsuccessful decoding, while

TABLE I

$A(x)$ AND $I(x)$ FOR THE $[8,4,4]$ REED-MULLER CODE AND $S(x)$ AND $D(x)$ FOR THE PARITY-CHECK MATRICES $\mathbf{H}_4$, $\mathbf{H}_5$, $\mathbf{H}_8$, $\mathbf{H}_{14}$, AND $\mathbf{H}^\star$.

	$A(x)$	$I(x)$
	$1+14x^4+x^8$	$14x^4+56x^5+28x^6+8x^7+x^8$

parity-check matrix	$S(x)$	$D(x)$
$\mathbf{H}_4$	$1+2x^3+24x^4+40x^5+28x^6+8x^7+x^8$	$2x^3+32x^4+56x^5+28x^6+8x^7+x^8$
$\mathbf{H}_5$	$1+18x^4+36x^5+28x^6+8x^7+x^8$	$18x^4+56x^5+28x^6+8x^7+x^8$
$\mathbf{H}_8$	$1+14x^4+24x^5+28x^6+8x^7+x^8$	$14x^4+56x^5+28x^6+8x^7+x^8$
$\mathbf{H}_{14}$	$1+14x^4+28x^6+8x^7+x^8$	$14x^4+56x^5+28x^6+8x^7+x^8$
$\mathbf{H}^\star$	$1+14x^4+28x^6+8x^7+x^8$	$14x^4+56x^5+28x^6+8x^7+x^8$

the iterative decoders based on $\mathbf{H}_4$ and $\mathbf{H}_5$ do not. Although $\mathbf{H}_{14}$ is larger than $\mathbf{H}_8$ and both achieve the maximum successful decoding probability, there are advantages in using $\mathbf{H}_{14}$ instead of $\mathbf{H}_8$. For instance, suppose that the erasure set is $\{1,2,3,7,8\}$. This erasure set is an incorrigible set since it contains $\{1,2,7,8\}$ which is the support of a non-zero codeword in the code. Therefore, any decoding method fails in retrieving the transmitted codeword. However, iterative decoding based on $\mathbf{H}_{14}$ succeeds in determining the erased bit c_3 from the parity-check equation $c_3+c_4+c_5+c_6=0$ since (00111100) is a row in $\mathbf{H}_{14}$. On the other hand, iterative decoding based on $\mathbf{H}_8$ does not succeed in retrieving any of the erased bits. Actually, since the coefficient of x^5 in the stopping set enumerator of $\mathbf{H}_{14}$ is zero, it follows that if the erasure set is a dead-end set of size five, it is always possible to retrieve one of the erased bits using $\mathbf{H}_{14}$. This is not true if matrix $\mathbf{H}_8$ is used instead.

We will show that the range $0 \le i \le 2$ specified by (10) for which $S_i = A_i$ can be considerably extended in case $\mathbf{H}$ is the complete parity-check matrix $\mathbf{H}^\star$. First we start with two lemmas.

Lemma 1: For any non-empty set $\mathcal{S} \subseteq \{1,2,\ldots,n\}$ which does not contain the support of a non-zero codeword, each binary vector of length $|\mathcal{S}|$ appears exactly $2^{n-k-|\mathcal{S}|}$ times as a row in $\mathbf{H}^\star_{\mathcal{S}}$.

Proof: Since $\mathcal{S}$ does not contain the support of a non-zero codeword, the $2^{n-k} \times |\mathcal{S}|$ matrix $\mathbf{H}^\star_{\mathcal{S}}$ has rank $|\mathcal{S}|$. Hence, there are $|\mathcal{S}|$ linearly independent rows in $\mathbf{H}^\star_{\mathcal{S}}$. The linear combinations of these $|\mathcal{S}|$ rows generate the space of all binary vectors of length $|\mathcal{S}|$, and thus each of these $2^{|\mathcal{S}|}$ vectors appears exactly $2^{n-k}/2^{|\mathcal{S}|}$ times as a row in $\mathbf{H}^\star_{\mathcal{S}}$. ■

Lemma 2: For any set $\mathcal{S} \subseteq \{1,2,\ldots,n\}$ which contains exactly one support $\mathcal{S}'$ of a non-zero codeword, each binary vector of length $|\mathcal{S}|$, with even weight on the positions indexed by $\mathcal{S}'$ and any weight on the positions indexed by $\mathcal{S} \setminus \mathcal{S}'$, appears exactly $2^{n-k-|\mathcal{S}|+1}$ times as a row in $\mathbf{H}^\star_{\mathcal{S}}$.

Proof: Since $\mathcal{S}$ contains exactly one support $\mathcal{S}'$ of a non-zero codeword, the $2^{n-k} \times |\mathcal{S}|$ matrix $\mathbf{H}^\star_{\mathcal{S}}$ has rank $|\mathcal{S}|-1$. Hence, there are $|\mathcal{S}|-1$ linearly independent rows in $\mathbf{H}^\star_{\mathcal{S}}$. The linear combinations of these $|\mathcal{S}|-1$ rows generate the space of all binary vectors of length $|\mathcal{S}|$ with even weight on the positions indexed by $\mathcal{S}'$ and any weight on the positions indexed by $\mathcal{S}\setminus\mathcal{S}'$, and thus each of these $2^{|\mathcal{S}|-1}$ vectors appears exactly $2^{n-k}/2^{|\mathcal{S}|-1}$ times as a row in $\mathbf{H}^\star_{\mathcal{S}}$. ■

Theorem 5: For any code,

$$S^\star_i = A_i \text{ for } i = 0,1,\ldots,\min\{\lceil 3d/2 \rceil - 1, n\}, \qquad (24)$$

i.e., the enumerators $S^\star(x)$ and $A(x)$ are equal in at least the first $\min\{\lceil 3d/2 \rceil, n+1\}$ coefficients.

Proof: Since the result is trivial for $i=0$, we may assume $1 \le i \le \min\{\lceil 3d/2 \rceil - 1, n\}$. Suppose that $\mathcal{S}$ is a stopping set of size i for $\mathbf{H}^\star$, which is not the support of a codeword. This set $\mathcal{S}$ contains at most one support of a non-zero codeword, since it follows from the Griesmer bound [10] that any linear code of dimension greater than 1 and Hamming distance at least d has a length of at least $d + \lceil d/2 \rceil = \lceil 3d/2 \rceil$. It follows from Lemmas 1 and 2 that $\mathbf{H}^\star_{\mathcal{S}}$ contains at least one row of weight one. Together with (6), we reach a contradiction to the assumption that $\mathcal{S}$ is a stopping set for $\mathbf{H}^\star$. Hence, any stopping set of size i for $\mathbf{H}^\star$ is the support of a codeword. In conclusion, $S^\star_i \le A_i$, and together with (18) we obtain the result presented in (24). ■

In the remainder of this section, we give a complete characterization of codes that have parity-check matrices for which $S(x) = A(x)$, i.e., codes with parity-check matrices for which every stopping set is a support of a codeword. For convenience, such codes are called minimum stopping. From (18), we conclude that a code is minimum stopping if and only if its optimal stopping set enumerator equals its weight enumerator, i.e., $S^\star(x) = A(x)$. We start by giving three classes of codes satisfying this condition.

(i) $\mathcal{R}_n$ is the $[n,1,n]$ repetition code consisting of the all-zero and all-one vectors of length n. From Theorem 5, it follows that $S^\star_i = A_i$ for $i = 0,1,\ldots,n$. Hence, $S^\star_0 = A_0 = 1$, $S^\star_n = A_n = 1$, and $S^\star_i = A_i = 0$ for $i = 1,2,\ldots,n-1$.

(ii) $\mathcal{F}_n$ is the $[n,n,1]$ full-code consisting of all binary vectors of length n. Clearly, $S^\star_i = A_i = \binom{n}{i}$ for $i = 0,1,\ldots,n$.

(iii) $\mathcal{Z}_n$ is the $[n,0,\infty]$ zero-code consisting of one codeword only, which is the all-zero vector of length n. Since all vectors of length n, including those of weight one, belong to the complete parity-check matrix of the code, it follows that $S^\star_i = A_i = 0$ for $i = 1,2,\ldots,n$, and $S^\star_0 = A_0 = 1$.

Next, we introduce a useful notation. For $i = 1,2,\ldots,t$, let $\mathcal{C}_i$ be a binary linear $[n_i,k_i,d_i]$ block code. Then, we define

$$\mathcal{C}_1 \oplus \mathcal{C}_2 \oplus \cdots \oplus \mathcal{C}_t \quad = \quad \{(\mathbf{c}_1,\mathbf{c}_2,\ldots,\mathbf{c}_t) : \mathbf{c}_i \in \mathcal{C}_i$$

$$\forall i = 1, 2, \ldots, t\}, \tag{25}$$

i.e., $\mathcal{C}_1 \oplus \mathcal{C}_2 \oplus \cdots \oplus \mathcal{C}_t$ is the set of all sequences obtained by juxtaposing codewords in $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_t$ in this order. Clearly, $\mathcal{C}_1 \oplus \mathcal{C}_2 \oplus \cdots \oplus \mathcal{C}_t$ is an $[n_1 + n_2 + \cdots + n_t, k_1 + k_2 + \cdots + k_t, \min\{d_1, d_2, \ldots, d_t\}]$ binary linear block code.

Finally, recall that two codes are equivalent if there is a fixed permutation of indices that maps one to the other [10], and that a code is said to have no zero-coordinates if and only if there is no index $i \in \{1, 2, \ldots, n\}$ such that $c_i = 0$ for every codeword $(c_1, c_2, \ldots, c_n)$.

Lemma 3: The code $\mathcal{C}_1 \oplus \mathcal{C}_2 \oplus \cdots \oplus \mathcal{C}_t$ is minimum stopping if and only if $\mathcal{C}_i$ is minimum stopping for all $i = 1, \ldots, t$.

Proof: First, notice that the code $\mathcal{C} = \mathcal{C}_1 \oplus \mathcal{C}_2 \oplus \cdots \oplus \mathcal{C}_t$ has a block diagonal parity check matrix $\mathbf{H}_1 \oplus \mathbf{H}_2 \oplus \cdots \oplus \mathbf{H}_t$ defined by

$$\mathbf{H}_1 \oplus \mathbf{H}_2 \oplus \cdots \oplus \mathbf{H}_t = \begin{pmatrix} \mathbf{H}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_t \end{pmatrix}, \tag{26}$$

where $\mathbf{H}_i$ is a parity check matrix for $\mathcal{C}_i$. The complete parity-check matrix $\mathbf{H}^\star$ of $\mathcal{C}$ has all elements of the row space of the matrix from (26) as its rows.

Next, let $\mathcal{S}$ be a subset of $\{1, 2, \ldots, n_1 + n_2 + \cdots + n_t\}$. For $i = 1, 2, \ldots, t$, define

$$\mathcal{S}_i = \{1 \leq m \leq n_i : m + \sum_{j=1}^{i-1} n_j \in \mathcal{S}\}. \tag{27}$$

Then, it follows from (25) that $\mathcal{S}$ is the support of a codeword in $\mathcal{C}_1 \oplus \mathcal{C}_2 \oplus \cdots \oplus \mathcal{C}_t$ if and only if $\mathcal{S}_i$ is the support of a codeword in $\mathcal{C}_i$ for all $i = 1, 2, \ldots, t$.

Further, $\mathcal{S}$ is a stopping set for $\mathbf{H}^\star$ if and only if $\mathcal{S}_i$ is a stopping set for $\mathbf{H}_i^\star$ for all $i = 1, 2, \ldots, t$. This follows from the fact that a sequence $\mathbf{r}$ is a row in $\mathbf{H}^\star$ if and only if it can be written as $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t)$ where $\mathbf{r}_i$ is a row in $\mathbf{H}_i^\star$ for all $i = 1, 2, \ldots, t$. Hence, a sequence $\mathbf{r}$ is a row in $\mathbf{H}_\mathcal{S}^\star$ if and only if it can be written as $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t)$ where $\mathbf{r}_i$ is a row in $\mathbf{H}_{i,\mathcal{S}_i}^\star$ for all $i = 1, 2, \ldots, t$. If for some $i = 1, 2, \ldots, t$, $\mathcal{S}_i$ is not a stopping set for $\mathbf{H}_i^\star$, then $\mathbf{H}_{i,\mathcal{S}_i}^\star$ has a row of weight one. Juxtaposing this row with the all-zero rows in $\mathbf{H}_{j,\mathcal{S}_j}^\star$, for all $j \neq i$, gives a row in $\mathbf{H}_\mathcal{S}^\star$ of weight one. This implies that $\mathcal{S}$ is not a stopping set for $\mathbf{H}^\star$. On the other hand, if for all $i = 1, 2, \ldots, t$, $\mathcal{S}_i$ is a stopping set for $\mathbf{H}_i^\star$, then for all i, $\mathbf{H}_{i,\mathcal{S}_i}^\star$ has no row of weight one. Juxtaposing rows of weights other than one yields a row of weight other than one. Hence, $\mathcal{S}$ is a stopping set for $\mathbf{H}^\star$.

We conclude that $\mathcal{S}$ is a stopping set for $\mathbf{H}^\star$ which is the support of a codeword in $\mathcal{C}_1 \oplus \mathcal{C}_2 \oplus \cdots \oplus \mathcal{C}_t$ if and only if, for all $i = 1, 2, \ldots, t$, $\mathcal{S}_i$ is a stopping set for $\mathbf{H}_i^\star$ which is the support of a codeword in $\mathcal{C}_i$. Hence, $\mathcal{C}_1 \oplus \mathcal{C}_2 \oplus \cdots \oplus \mathcal{C}_t$ is minimum stopping if and only if, for all $i = 1, 2, \ldots, t$, $\mathcal{C}_i$ is minimum stopping. ■

Lemma 4: Let $\mathcal{C}$ be a minimum stopping binary linear $[n, k, d]$ block code with $d \geq 2$ and no zero-coordinates. If $d = n$, then $\mathcal{C}$ is $\mathcal{R}_n$. Otherwise, $k \geq 2$ and $\mathcal{C}$ is equivalent to $\mathcal{R}_d \oplus \mathcal{C}''$ for some binary linear $[n - d, k - 1, d'']$ block code $\mathcal{C}''$ with $d'' \geq 2$ and no zero-coordinates.

Proof: Up to equivalence, we may assume that $\mathcal{C}$ has a codeword composed of d ones followed by $n - d$ zeros. In particular, the first d columns in any given parity check matrix of $\mathcal{C}$ are linearly dependent, but no $d-1$ columns are such. The row space of the submatrix composed of these first d columns has dimension $d - 1$ and a sequence of length d belongs to this row space if and only if its weight is even. Therefore, in case $d = n$, $\mathcal{C}$ has a full-rank $(d - 1) \times d$ parity check matrix of the form

$$\mathbf{H} = \mathbf{H}_d'' = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix}, \tag{28}$$

which shows that $\mathcal{C}$ is the repetition code of length n. Further, in case $d \leq n - 1$, $\mathcal{C}$ has a full-rank $(n - k) \times n$ parity check matrix of the form

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_d'' & \mathbf{H}' \\ \mathbf{0} & \mathbf{H}'' \end{pmatrix}, \tag{29}$$

where $\mathbf{H}'$ and $\mathbf{H}''$ are $(d - 1) \times (n - d)$ and $(n - k - d + 1) \times (n - d)$ matrices, respectively. Notice that $\mathbf{H}''$ has at least one row since $d \leq n - k + 1$ by the Singleton bound and if equality holds with $2 \leq d \leq n - 1$, then $\mathcal{C}$ is the even weight code of length $n \geq 3$ which has $\{1, 2, 3\}$ as a stopping set for $\mathbf{H}^\star$ which is not the support of a codeword. This contradicts the assumption that $\mathcal{C}$ is minimum stopping. Clearly, $\mathbf{H}''$ is a matrix of rank $n - k - d + 1$ since $\mathbf{H}$ in (29) is a full-rank matrix. If $k = 1$, then $\mathbf{H}''$ has rank $n - d$ and $\mathcal{C}$ has zero-coordinates. Therefore, $k \geq 2$. To complete the proof, it suffices to show that the row space of $\mathbf{H}'$ is a subspace of the row space of $\mathbf{H}''$ since, in this case, by elementary row operations, $\mathcal{C}$ has a parity-check matrix

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_d'' & \mathbf{0} \\ \mathbf{0} & \mathbf{H}'' \end{pmatrix}, \tag{30}$$

and thus $\mathcal{C} = \mathcal{R}_d \oplus \mathcal{C}''$ where $\mathcal{C}''$ is the code with parity check matrix $\mathbf{H}''$. This code has length $n - d$, dimension $k - 1$, and Hamming distance $d'' \geq d \geq 2$ with no zero-coordinates. Now, suppose, to get a contradiction, that the above is not true, i.e., the row space of $\mathbf{H}'$ is not a subspace of the row space of $\mathbf{H}''$. Then, the null space of $\mathbf{H}''$ is not a subspace of the null space of $\mathbf{H}'$. Let $\mathbf{c}''$ be a vector of length $n - d$ which belongs to the null space of $\mathbf{H}''$ but not to the null space of $\mathbf{H}'$. Up to equivalence, we may assume that $\mathbf{c}''$ is composed of w ones followed by $n - d - w$ zeros, where w, $1 \leq w \leq n - d$, is the weight of $\mathbf{c}''$. We claim that $\{1, 2, \ldots, d + w\}$ is a stopping set for $\mathbf{H}^\star$ which is not the support of a codeword in $\mathcal{C}$. From (29), we have

$$\mathbf{H}_{\{1,2,\ldots,d+w\}} = \begin{pmatrix} \mathbf{H}_d'' & \mathbf{H}'_{\{1,2,\ldots,w\}} \\ \mathbf{0} & \mathbf{H}''_{\{1,2,\ldots,w\}} \end{pmatrix}. \tag{31}$$

Notice that any nontrivial linear combination of the rows of $\mathbf{H}_d''$ yields a non-zero vector of even weight. Furthermore, since $\mathbf{c}''$, which starts with w ones followed by $n - d - w$ zeros, is in the null space of $\mathbf{H}''$, it follows that any linear combination of the rows of $\mathbf{H}''_{\{1,2,\ldots,w\}}$ yields an even weight vector. We conclude that no linear combination of the rows of $\mathbf{H}_{\{1,2,\ldots,d+w\}}$ yields a vector of weight one. Hence, $\{1,2,\ldots,d+w\}$ is a stopping set for $\mathbf{H}^\star$. Next, notice that if $\{1,2,\ldots,d+w\}$ is the support of a codeword in $\mathcal{C}$, then the columns in $\mathbf{H}_{\{1,2,\ldots,d+w\}}$ in (31) should add up to zero. From (28), we know that the first d columns add up to zero. Therefore, the columns of $\mathbf{H}'_{\{1,2,\ldots,w\}}$ should add up to the zero. However, this cannot be the case as $\mathbf{c}''$, which starts with w ones followed by $n-d-w$ zeros, is not in the null space of $\mathbf{H}'$. In conclusion, we have shown that $\{1,2,\ldots,d+w\}$ is a stopping set for $\mathbf{H}^\star$ which is not the support of a codeword in $\mathcal{C}$. This contradicts the fact that $\mathcal{C}$ is minimum stopping. ■

Lemma 5: If $\mathcal{C}$ is a minimum stopping binary linear $[n,k,d]$ block code with $d \geq 2$ and no zero-coordinates, then $\mathcal{C}$ is equivalent to

$$\mathcal{R}_{n_1} \oplus \mathcal{R}_{n_2} \oplus \cdots \oplus \mathcal{R}_{n_t}, \tag{32}$$

for some integers $n_1, n_2, \ldots, n_t \geq 2$ and $t \geq 1$ such that $n_1 + n_2 + \cdots + n_t = n$.

Proof: The lemma trivially holds for all codes of lengths two or less. We use induction and assume that it holds for all codes of length less than n. From Lemma 4, we know that either $\mathcal{C}$ is equivalent to $\mathcal{R}_n$, which is consistent with the statement of the lemma, or $k \geq 2$ and $\mathcal{C}$ is equivalent to $\mathcal{R}_d \oplus \mathcal{C}''$ for some binary linear $[n-d, k-1, d'']$ block code $\mathcal{C}''$ with $d'' \geq 2$ and no zero-coordinates. From Lemma 3, we know that $\mathcal{C}''$ is a minimum stopping code. Since $\mathcal{C}''$ has length $n-d < n$, it follows from the induction hypothesis that $\mathcal{C}''$ is equivalent to $\mathcal{R}_{m_1} \oplus \cdots \oplus \mathcal{R}_{m_v}$, for some integers $m_1, \ldots, m_v \geq 2$ and $v \geq 1$ such that $m_1 + \cdots + m_v = n-d$. Then, $\mathcal{R}_d \oplus \mathcal{C}''$ has the same form as given in the lemma. ■

Theorem 6: A binary linear $[n,k,d]$ block code $\mathcal{C}$ is minimum stopping, i.e., satisfies $S^\star(x) = A(x)$, if and only if it is equivalent to

$$\mathcal{R}_{n_1} \oplus \mathcal{R}_{n_2} \oplus \cdots \oplus \mathcal{R}_{n_u} \oplus \mathcal{F}_{n_F} \oplus \mathcal{Z}_{n_Z}, \tag{33}$$

for some nonnegative integers $n_1, n_2, \ldots, n_u, n_F, n_Z$ and u, where $n_1, n_2, \ldots, n_u \geq 2$ and $n_1 + n_2 + \cdots + n_u + n_F + n_Z = n$.

Note: In the theorem, we allow $u = 0$ in which case $\mathcal{C}$ is equivalent to $\mathcal{F}_{n_F} \oplus \mathcal{Z}_{n_Z}$. We also allow $n_F = 0$ and/or $n_Z = 0$, in which case the corresponding code with length zero disappears from $\mathcal{R}_{n_1} \oplus \mathcal{R}_{n_2} \oplus \cdots \oplus \mathcal{R}_{n_u} \oplus \mathcal{F}_{n_F} \oplus \mathcal{Z}_{n_Z}$.

Proof: The "if"-part of the theorem follows from Lemma 3 and the observations that the $S^\star(x) = A(x)$ property holds for any repetition code $\mathcal{R}_{n_i}$, the full-code $\mathcal{F}_{n_F}$, and the zero-code $\mathcal{Z}_{n_Z}$.

Next, we proof the "only if"-part of the theorem. Up to equivalence, we may assume that $\mathcal{C} = \mathcal{C}' \oplus \mathcal{F}_{n_F} \oplus \mathcal{Z}_{n_Z}$, where n_Z is the number of zero-coordinates of $\mathcal{C}$, n_F is the number of codewords of weight one in $\mathcal{C}$, i.e., the number of all-zero columns in any parity check matrix of $\mathcal{C}$, and $\mathcal{C}'$ is a binary linear code of length $n - n_F - n_Z$ with $d \geq 2$ and no zero-coordinates. Here we assume that if $n - n_F - n_Z$, n_F, or n_Z equal zero, then the corresponding code disappears from $\mathcal{C}' \oplus \mathcal{F}_{n_F} \oplus \mathcal{Z}_{n_Z}$. If $\mathcal{C}'$ does not disappear, then it can be written as stated in Lemma 5. ■

V. Conclusion

In this paper, we examined how the performance of iterative decoding when applied to a binary linear block code over an erasure channel depends on the parity-check matrix representing the code. This code representation determines the complexity of the decoder. We have shown that there is a trade-off between performance and complexity.

In particular, we have shown that, regardless of the choice of the parity-check matrix, the stopping set enumerator differs from the weight enumerator except for a degenerate class of codes. In spite of that, it is always possible to choose parity-check matrices for which the dead-end set enumerator equals the incorrigible set enumerator. Iterative decoding based on such matrices is optimal, in the sense that it gives the same probability of unsuccessful decoding on the binary erasure channel as an exhaustive decoder. We presented bounds on the number of rows in parity-check matrices with optimal dead-end set enumerators, thus bounding the complexity of iterative decoding achieving optimal performance.

References

[1] K.A.S. Abdel-Ghaffar and J.H. Weber. (2006, March). Complete enumeration of stopping sets of full-rank parity-check matrices of Hamming codes. arXiv:cs.IT/0603007. [Online].
Available: http://www.arxiv.org/list/cs.IT/0603

[2] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1115–1131, June 2004.

[3] C. Di, D. Proietti, I.E. Telatar, T.J. Richardson, and R.L. Urbanke, "Finite-length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1570-1579, June 2002.

[4] J. Han and P. H. Siegel. (2005, November). Improved upper bounds on stopping redundancy. arXiv:cs.IT/0511056. [Online].
Available: http://www.arxiv.org/list/cs.IT/0511

[5] H. D. L. Hollmann and L. M. G. M. Tolhuizen. (2005, July). On parity check collections for iterative erasure decoding that correct all correctable erasure patterns of a given size. arXiv:cs.IT/0507068. [Online].
Available: http://www.arxiv.org/list/cs.IT/0507

[6] R. Ikegaya, K. Kasai, T. Shibuya, and K. Sakaniwa, "Asymptotic weight and stopping set distributions for detailedly represented irregular LDPC code ensembles," Proceedings of the IEEE International Symposium on Information Theory, Chicago, USA, p. 208, June 27 - July 2, 2004.

[7] N. Kashyap and A. Vardy, "Stopping sets in codes from designs," Proceedings of the IEEE International Symposium on Information Theory, Yokohama, Japan, p. 122, June 29 - July 4, 2003.

[8] C. Kelley, D. Sridhara, J. Xu, and J. Rosenthal, "Pseudocodeword weights and stopping sets," Proceedings of the IEEE International Symposium on Information Theory, Chicago, USA, p. 67, June 27 - July 2, 2004.

[9] M.G. Luby, M. Mitzenbacher, M.A. Shokrollahi, and D.A. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 569–584, February 2001.

[10] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.

[11] A. Orlitsky, R. Urbanke, K. Viswanathan, and J. Zhang, "Stopping sets and the girth of Tanner graphs," Proceedings of the IEEE International Symposium on Information Theory, Lausanne, Switzerland, p. 2, June 30 - July 5, 2002.

[12] A. Orlitsky, K. Viswanathan, and J. Zhang, "Stopping set distribution of LDPC code ensembles," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 929–953, March 2005.

[13] M. Schwartz and A. Vardy, "On the stopping distance and the stopping redundancy of codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 922–932, March 2006.

[14] J. H. Weber and K. A. S. Abdel-Ghaffar, "Stopping set analysis for Hamming codes," Proceedings of the Information Theory Workshop on Coding and Complexity, Rotorua, New Zealand, pp. 244–247, August 28–September 1, 2005.

[15] S. -T. Xia and F. -W. Fu, “On the stopping distance of finite geometry LDPC codes,” *IEEE Commun. Letters*, vol. 10, no. 5, pp. 381–383, May 2006.